Title
Large Phenotypic Enhancement of Structured Random RNA Pools.

Journal
Journal of the American Chemical Society, 142(4)

ISSN
0002-7863

Authors
Chizzolini, Fabio
Passalacqua, Luiz FM
Oumais, Mona
et al.

Publication Date
2020

DOI
10.1021/jacs.9b11396

Peer reviewed

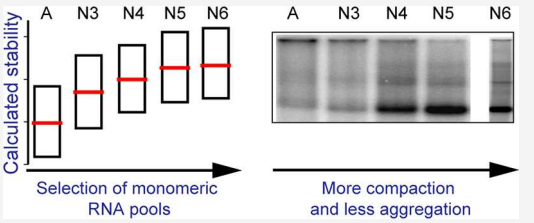
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Large Phenotypic Enhancement of Structured Random RNA Pools

Fabio Chizzolini,[#] Luiz F. M. Passalacqua,[#] Mona Oumais,[#] Armine I. Dingilian, Jack W. Szostak,* and Andrej Lupták*

Cite This: https://dx.doi.org/10.1021/jacs.9b11396
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Laboratory evolution of functional RNAs has applications in many areas of chemical and synthetic biology. In vitro selections critically depend on the presence of functional molecules, such as aptamers and ribozymes, in the starting sequence pools. For selection of novel functions the pools are typically transcribed from random-sequence DNA templates, yielding a highly diverse set of RNAs that contain a multitude of folds and biochemical activities. The phenotypic potential, the frequency of functional RNAs, is very low, requiring large complexity of starting pools, surpassing 10^{15} different sequences, to identify highly active isolates. Furthermore, the majority of random sequences is not structured and has a high propensity for aggregation; the in vitro selection process thus involves not just enrichment of functional RNAs, but also their purification from aggregation-prone "free-riders". We reasoned that purification of the nonaggregating, monomeric subpopulation of a random-sequence RNA pool will yield pools of folded, functional RNAs. We performed six rounds of selection for monomeric sequences and show that the enriched population is compactly folded. In vitro selections originating from various mixtures of the compact pool and a fully random pool showed that sequences from the compact pool always dominate the population once a biochemical activity is detectable. A head-to-head competition of the two pools starting from a low (5×10^{12}) sequence diversity revealed that the phenotypic potential of the compact pool is about 1000-times higher than the fully random pool. A selection for folded and monomeric RNA pools thus greatly increases the frequency of functional RNAs from that seen in random-sequence pools, providing a facile experimental approach to isolation of highly active functional RNAs from low-diversity populations.



INTRODUCTION

Functional RNAs fold into specific structures that form binding sites for aptamer ligands and active sites of ribozymes.[1−4] These structures range from simple stem-loops and G-quadruplexes to multihelical junctions, pseudoknots, and their combinations, giving rise to intricate folds that endow these RNAs with their biochemical activities. The RNAs typically fold hierarchically, using base-paired regions as structural modules, and interspersed single-stranded elements to make tertiary interactions, functional sites, and peripheral loops.[5,6]

In vitro selection experiments harness the enormous potential of molecular evolution to yield functional RNAs.[7] The key for the success of any selection is the ability to sample a sufficient number of sequences and structures that are diverse enough to allow isolation of macromolecules carrying the desired phenotype. High-complexity DNA pools, typically composed of random sequences flanked by sequence-invariant primer-binding regions, are used as templates for in vitro transcription and experimental isolation of desired functional RNAs.[8−10] Because at least a weak activity has to be present in the starting pool for a selection to yield a functional RNA, the composition of the starting population is critical for the success of the entire selection experiment.[11] The phenotypic potential of the starting pool, the frequency of functional sequences, is thus a key factor of the experiment, but this property has been difficult to modulate experimentally.

A dominant property of random-sequence pools is their concentration-dependent aggregation, which potentially obscures active sequences. This propensity for aggregation may be the result of individual (unevolved) sequences to assume multiple competing conformations,[12] particularly in the presence of Mg^{2+} or other multivalent cations that support the formation of tertiary and quaternary structures, and thus allow nonfunctional "free-riders" to survive the selection process. While longer and more diverse random regions are desirable to increase the chance of isolating complex RNA folds[13] in the context of a randomized library, increasing the length of a random pool has been shown to lower the yield of small aptamers.[14] To address this issue, two approaches have previously been used to increase the structured fraction of the starting pool and yield a higher frequency of functional sequences: (i) incorporation of a structured domain next to or within the random sequence[15−17] and (ii) biasing the nucleotide composition of the random region to increase the

Received: October 23, 2019
Published: December 30, 2019

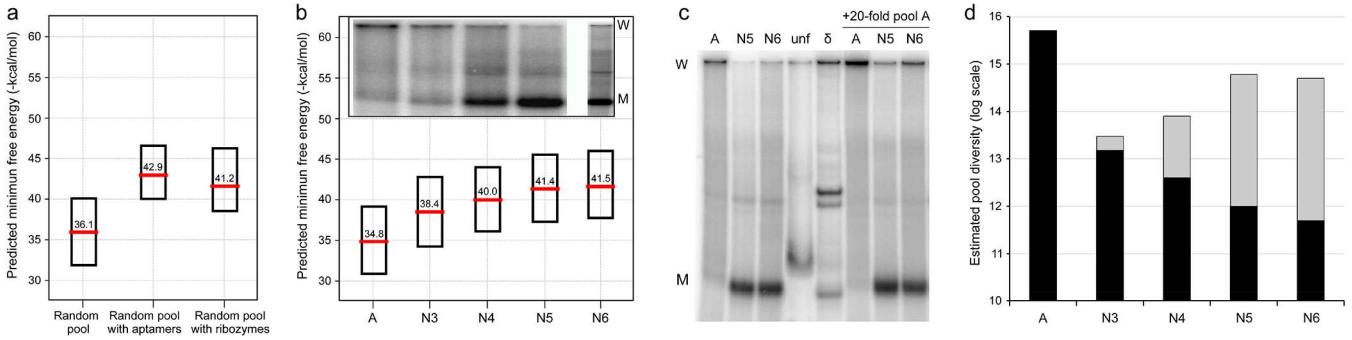


Figure 1. Folding propensities of random-sequence and monomeric RNA pools. (a) A plot of predicted minimum free energy of folding (MFE) of one million sequences of in silico, generated RNA pools consisting of either a fully random segment or functional motifs flanked by random regions. The functional motifs consisted of pooled aptamers or ribozymes. Red bar shows the average MFE for each pool and the boxes indicate one quartile deviation for each distribution. MFE distributions, including the sequences with the ten most extreme MFEs, of individual aptamer and ribozyme pools are shown in Figure S1a. RNA pools containing functional motifs show significantly higher stability, suggesting that the process of in vitro selection shifts the population toward more folded functional molecules. (b) In vitro evolution of a random RNA pool (A) from a population of aggregation-prone sequences of low average stability into a population of monomeric RNAs results in gradual increase of average stability, which plateaus after five rounds of enrichment (pools N3−N6). MFE analysis of sequences from the starting random-sequence DNA pool (A) and DNAs derived from the monomeric RNA populations selected using native PAGE. One million sequences were chosen randomly from an HTS of each pool. Red bar shows the average MFE for each pool, and the boxes indicate one quartile deviation for each distribution. Native PAGE analysis (inset; 10% polyacrylamide containing 0.5 × TBE and 10 mM $Mg(OAc)_2$) shows that the pools become less aggregated, moving the majority of the RNAs from the proximity of the starting point of the PAGE (W, well) to a population migrating as monomeric RNAs (M). This large mobility increase is retained even at elevated pool concentrations (Figure S1b). (c) Native PAGE analysis of the starting random pool (A) and the monomeric pools (N5, N6) shows that the monomeric pools migrate significantly faster than an unfolded RNA (unf) and near a genomic HDV ribozyme (δ; both of the same length as the pools), which represents a highly compact functional RNA. Incubation of the pools with a 20-fold excess of unlabeled pool A (right side of the gel) results in further aggregation, whereas the majority of the pools N5 and N6 is resistant to aggregation. (d) Graph of estimated maximum complexities of individual pools. Black bars correspond to the sequences derived from pool A. Gray bars show the additional sequence diversity introduced through mutagenic PCR. The diversity values assume perfect recovery of sequences from native PAGE gels and unbiased amplification. The native-PAGE selection decreased the fraction of the starting pool (A) by ∼4 orders of magnitude and the mutagenic PCR.

prevalence of base-paired elements within the pool.[18,19] However, both of these approaches are biased toward the built-in structural features and cannot sample a full range of RNA topologies and folds, presumably limiting their phenotypic potential.

We developed an experimental approach that decreases the aggregation of a random-sequence pool and increases its phenotypic potential. We show that isolation of the monomeric fraction of a random-sequence pool yields a population enriched in compact, structured RNAs. In vitro selections comparing this compact pool with a fully random pool of the same length show that the compact pool is as rich in functional RNAs as a fully random pool, but because the sequence diversity of the compact pool is significantly lower, the frequency of functional RNAs is greatly increased. Thus, a preselection for a folded, monomeric pool provides an experimental approach to increase the phenotypic potential of random-sequence libraries.

RESULTS

Functional RNAs Are More Structured than Random RNAs. Previous work comparing several folded RNAs with randomly permuted sequences of similar length showed that the nonevolved sequences can assume compact folds and specific secondary structures; however, the analysis was performed with only three functional RNAs and 20 randomized sequences, most of which exhibited structural heterogeneity on nondenaturing polyacrylamide gel electrophoresis (PAGE) gels.[12] To provide a broader comparison of functional and random sequences and to address the question of whether structured pools would be enriched in functional sequences relative to fully random pools, we first generated in silico several pools of random RNAs of identical length but different nucleotide composition, each with 1 million sequences. The first RNA pool was designed to mimic a pool of nonevolved RNAs, composed of a random stretch of 70 nucleotides (nts) flanked by 20-nt primer-binding sequences. Next, we designed ten different pools, each mimicking a population of potentially functional RNAs. The first three pools had functional motifs consisting of well-characterized RNA aptamers, flanked by two random regions that extended the aptamer sequences to 70 nts, flanked again by the same two constant regions as in the fully random library. The three pools contained an ATP-binding aptamer (40 nts),[20] a class I GTP-binding aptamer (41 nts),[15,21] and the Malachite Green aptamer (38 nts).[22] We also generated seven RNA pools with functional regions consisting of well-characterized ribozymes: a class I ligase ribozyme (96 nts),[23] an aminoacylase ribozyme (90 nts),[24] a kinase ribozyme (89 nts),[25] the R3C ligase ribozyme (73 nts),[26] the L1X6c ligase ribozyme (71 nts),[27,28] a class III ligase ribozyme (56 nts),[23] and a class II ligase ribozyme (87 nts).[23,29] Because some of these ribozymes are larger than the aptamers, the functional domains were only flanked by two random regions, without any additional constant regions. The total length of each RNA sequence contained in all of the RNA pools was kept constant. Next, we calculated the predicted minimum free energy (MFE) of folding for each RNA sequence in the pools, and plotted the resulting distributions of MFEs to compare the three aptamer-containing pools and the seven ribozyme-containing pools with the fully random pool (Figure 1a; see Figure S1 for the MFE distributions of individual aptamer and ribozyme pools).

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Compared to the fully random pool, the average predicted MFE of both the aptamer and ribozyme pools was significantly shifted toward more structured RNAs (Figures 1a and S1a). Furthermore, the least stable tails of the distributions, representing the sequences with the lowest predicted secondary structure content, differed even more between the random and the functional pools, with the ten least-structured sequences from the random pool mapping significantly below the bottom ten sequences from any of the functional pools (Figure S1a). On the other hand, the ten most stable sequences from the random pool had MFEs similar to the most stable sequences of the aptamer and ribozyme pools (Figure S1). These calculations showed that a random-sequence pool is on average less folded and has a broader distribution of secondary structures than pools containing functional domains. The results provided a computational foundation for our hypothesis that a random RNA pool preselected for nonaggregating, monomeric sequences contains a higher proportion of structured RNA molecules, and that these folded populations are enriched for biochemically active motifs, increasing the phenotypic potential of random libraries.

Selection of Monomeric Subpopulation from a Random RNA Pool Yields Aggregation-Resistant Sequences. To test the hypothesis that the monomeric fraction of a random-sequence pool is more folded and has a higher phenotypic potential, we purified nonaggregating, monomeric sequences from a random RNA pool. First, a large-scale transcription reaction of a random-sequence DNA pool (pool A, containing 70 nts of random sequence flanked by two primer-binding regions to yield 110-nt RNA transcripts) was fractionated using a PAGE gel composed of a denaturing top section containing urea, followed by a nondenaturing (native) section, lacking urea and containing Mg^{2+} (see Material and Methods for details). By isolating only the fastest migrating RNAs from the nondenaturing part of the gel, we reduced the overall diversity of the pool, presumably eliminating aggregation-prone molecules retained either in the well or in the denaturing portion of the gel, while increasing the prevalence of monomeric species. This purification technique was applied under progressively less denaturing conditions for a total of three times before the first reverse-transcription and amplification were performed. Having undergone three native-PAGE purifications, this pool was termed N3. To distinguish the selected pool from the starting population, we introduced a point mutation in the 3′ constant region, changing a StyI restriction endonuclease binding site to that of the MnlI enzyme (see Table S1 for all relevant DNA sequences). The mutation also shifts the restriction-enzyme cut site such that a mixed pool, containing both pool-A (fully random) and pool-N (monomeric pools) sequences, could be digested by these two restriction enzymes to yield distinct bands on analytical agarose or PAGE gels, allowing a straightforward measurement of the relative proportion of the two pools.

The same native-PAGE purification was performed three more times, with reverse transcription and PCR amplification after each PAGE fractionation, yielding pools N4, N5, and N6. Furthermore, low-level (0.0076 mutations/nt/5 cycles of PCR) mutagenesis[30,31] was applied to all DNA pools (N3–N6) to increase the sequence diversity near (within a few mutations) the selected sequences. As expected, native-PAGE analysis of transcripts from the naïve (A) and selected pools (N3, N4, N5, and N6) showed that the population shifted from predominantly aggregated RNAs, immobilized near the well, to RNAs migrating predominantly as a monomeric band (Figure 1b inset). The native-PAGE analysis of the pools also showed that the monomeric pools migrated about as fast as the self-cleaved form of the genomic HDV ribozyme, a stably folded functional RNA of the same length as the pools.[32] This PAGE mobility was noticeably faster than that of a construct that does not fold into a stable structure and was used as an unfolded-control RNA of the same covalent length (Figures 1c and S1b).

We next asked whether the selection yielded monomeric RNAs because they are simply mutually nonaggregating or whether they are also resistant to aggregation-prone random sequences. As expected, analysis of the selected pools showed higher aggregation of pools A and N3 at higher concentrations (~20 vs ~10 μM), whereas pools N4 and N5 remained monomeric and migrated as folded RNAs at both concentrations, suggesting that they are mutually nonaggregating (Figure S1b). We also transcribed pools A, N5, and N6 independently and analyzed them in mixtures with 20-fold excess of unlabeled pool A to assess the resistance to aggregation with fully random sequences. All three pools (A, N5, and N6) exhibited more aggregation in the presence of excess of pool A; however, for pool A, this shift of the population was dramatic, whereas for the selected pools (N5 and 6), it was minimal (Figure 1c). These results demonstrate that the preselection for the monomeric subpopulation of a random pool yields sequences that are not only nonaggregating among themselves, but also resistant to aggregation with fully random sequences, suggesting that the selection of the monomeric subpopulation of the pool enriched for tightly and uniquely folding sequences.

Given the results obtained by native PAGE analysis (Figures 1b,c and S1b), we also expected that the selected pools contain more secondary structure elements than those of pool A. As a proxy measurement of secondary structure, we performed high-throughput sequencing (HTS) of the naïve (A) and selected pools (N3, N4, N5, and N6), and calculated the predicted MFEs of 1 million sequences from each pool. The MFE distributions showed that the average MFE of the pool A was significantly lower than for the selected pools, while the predicted stability increased with selection round, reaching a plateau for pools N5 and N6 (Figure 1b). Notably, the average MFE of the in silico, generated random pool and the sequenced random pool A had almost identical average MFEs, whereas the MFEs of the selected pools, particularly N5 and N6, were in strong accordance with the average MFEs of the in silico, generated pools containing functional domains (Figure 1a,b). These results suggested that not only was the selected subpopulation more compact and structured, but it was likely enriched for functional RNAs, as well.

To estimate the sequence diversity of the selected pools, we measured the fraction of RNA recovered from each native-PAGE step and calculated the fraction of the pool A sequences remaining in each pool N3–N6 (Figure 1d and Table 1). Assuming faithful and unbiased amplification of each recovered sequence, and accounting for the additional diversity introduced by the mutagenic PCR, we also estimated the maximum complexity of each pool. Given that the starting diversity of the pool A (5×10^{15}) undersamples the theoretical complexity ($4^{70} \approx 10^{42}$) by ~26 orders of magnitude, the sequences sampled in this pool were highly sparse in the total sequence space. Any sequence diversity introduced by the mutagenic PCR in addition to the pool-A–derived sequences

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX–XXX

Table 1. Estimated Complexity and Composition of in Vitro Selected Pools

pool	maximum diversity	maximum fraction of pool A
A	5×10^{15}	1
N3	$\lesssim 3 \times 10^{13}$	$\lesssim 0.003$
N4[a]	$\lesssim 8 \times 10^{13}$	$\lesssim 0.0008$
N5[a]	$\lesssim 6 \times 10^{14}$	$\lesssim 0.0002$
N6[a]	$\lesssim 5 \times 10^{14}$	$\lesssim 0.0001 \lesssim 5 \times 10^{11}$ pool A molecules remaining

[a]These pools have been amplified using mutagenic PCR, which increases the theoretical maximum sequence diversity beyond the sequences originating from pool A.

($\sim 10^3$-fold in pool N6) then represents local sequence variation around the monomeric, structured sequences enriched during the native-PAGE selection process, and not broad sampling of the theoretical sequence space. The mutagenic PCR is also a mechanism for potential increase of the number of functional RNAs beyond the sequences directly inherited from pool A. For the final pool, N6, the real complexity likely falls within a range defined by remaining pool A sequences ($\sim 5 \times 10^{11}$) and the additional diversity resulting from mutagenesis introduced during the amplification steps (total estimated diversity $\lesssim 5 \times 10^{14}$). Moreover, because some sequences are likely suppressed during transcription, reverse transcription, and PCR, the real molecular diversity of the N pools is likely somewhat lower than the estimated complexity range. We do not have an experimental approach to measure these biases; however, HTS analysis of the selected pools did not reveal any gross sequence bias that would result from the selection and amplification steps.

To further characterize the monomeric pools, we performed several experiments designed to establish whether the monomer selection led to an increase in the fraction of structured regions. We first used S1 nuclease, which preferentially cleaves single-stranded regions of RNAs,[33] to measure the susceptibility of the pools to RNase degradation. We incubated pools A, N5, and N6 with the nuclease and analyzed the populations at various time-points using high-resolution denaturing PAGE. We observed that, for the pool A, the bands corresponding to the full-length RNAs disappeared faster and to a greater extent than for the selected pools N5 and N6 (Figure 2a). Furthermore, the intensity of the fragments corresponding to RNAs between 66 and 89 nts long, presumably representing longer stretches of nuclease-inaccessible secondary or tertiary structures, was also higher in the selected pools (Figure 2a). A similar effect was observed for RNA sizes of about 45 nts, whereas the opposite was observed for very short RNA sequences of 11−13 nts (Figure 2a). These results showed that the monomeric pools are enriched for more nuclease-resistant, structured RNAs.

Next, as an independent experimental approach for assessing the secondary structure content of the pools, we performed a titration of the intercalating dye ethidium bromide. Ethidium fluorescence greatly increases upon intercalation in double-stranded regions of nucleic acids.[34] If the process of monomer enrichment of the selected pools led to an increase in the prevalence of double-stranded regions, the fluorescence of the intercalating dye should increase as well. Titration of ethidium bromide into purified pools A, N5, and N6 showed a significant increase in fluorescence for the selected pools,

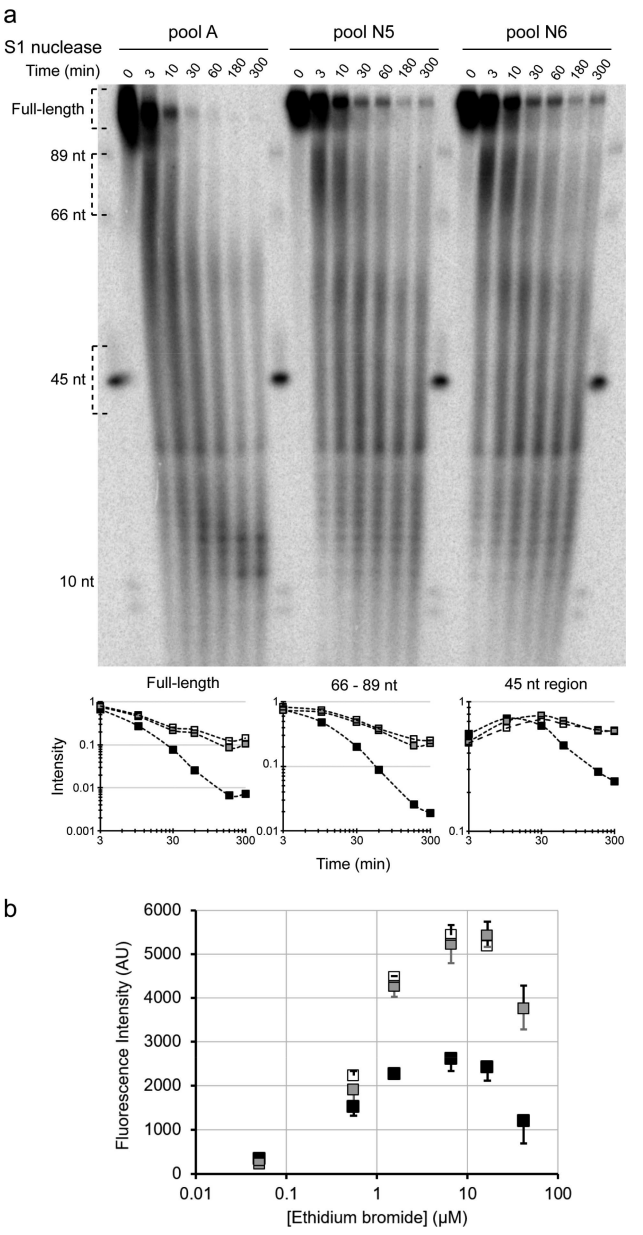


Figure 2. Biochemical characterization of the RNA pools. (a) S1 nuclease probing of pools A, N5, and N6. High-resolution PAGE analysis of the S1 digestion kinetics shows faster digestion of the fully random pool A than the monomeric pools N5 and N6, which appear to degrade to relatively stable domains above the 66-nt marker, as well as populations of shorter RNA domains, indicating that the monomeric pools are, on average, more tightly folded. The graphs bellow the gel image show the time courses of full-length, 66−89-nt, and 45-nt segments normalized to the starting material. The analysis reveals faster S1 degradation of pool A than N5 or N6. Accumulation of very short (11−13 nts) products follows the opposite trend. (b) Fluorescence due to ethidium intercalation into the paired regions of the pools A (black squares), N5 (gray), and N6 (open). Average fluorescence intensities for duplicate experiments are shown for each ethidium titration, indicating that the N5 and N6 pools contain about twice as much secondary structure elements as the pool A. Error bars represent the average deviation of the two measurements.

compared to pool A (Figure 2b), again indicating a higher prevalence of base-paired regions in the selected pools.

Lastly, we performed 2′ acylation of the RNA backbone.[35] Since the acylation reaction preferentially targets solvent-

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

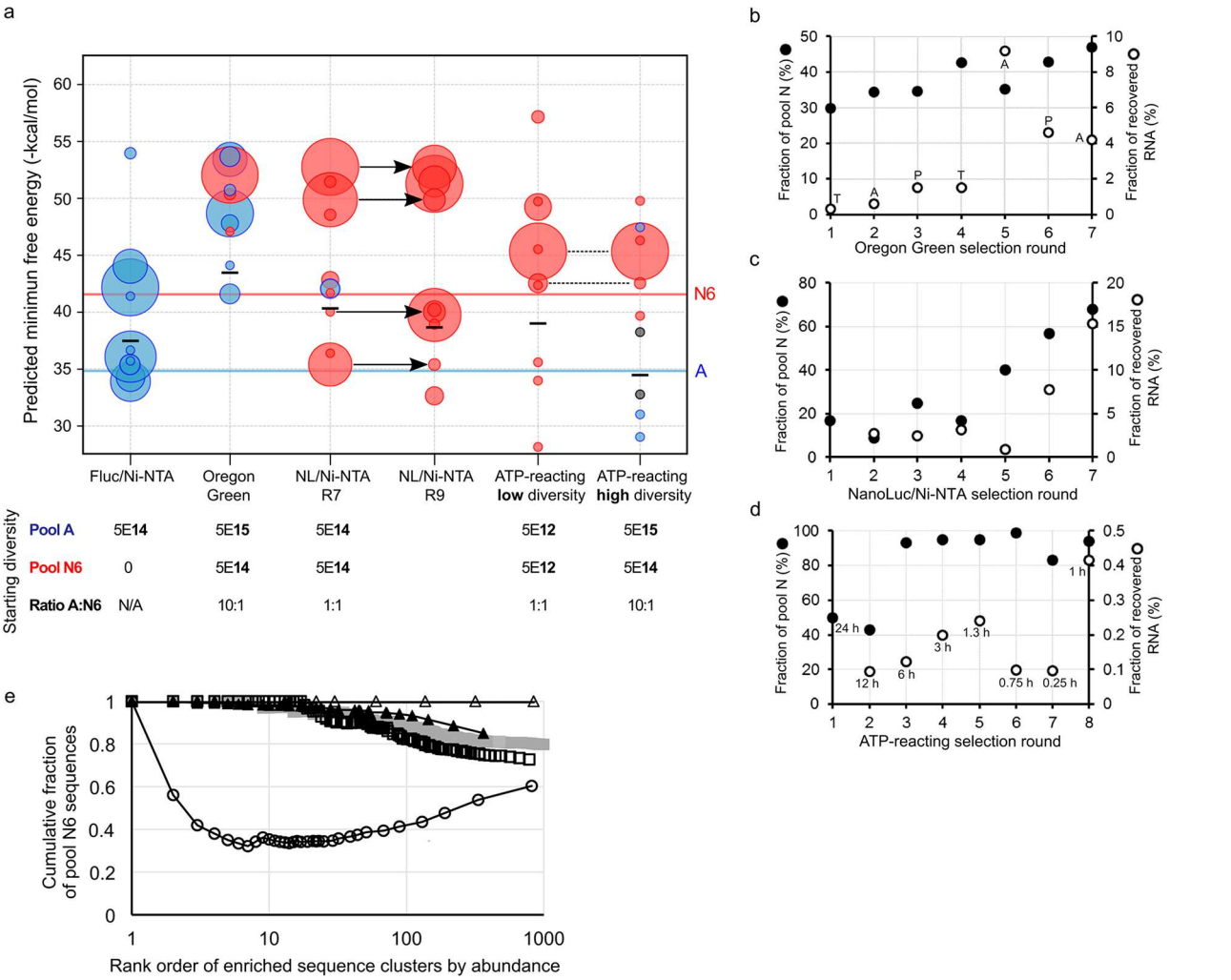


Figure 3. In vitro selections from different compositions of starting pools. (a) Predicted MFEs of the ten most abundant sequences for each of the in vitro selection experiments. The relative abundance of each sequence indicated by the circle diameter; with blue and red circles representing sequences originating from the pools A and N6, respectively. Gray circles represent sequences which could not be assigned to either of the two pools due to mutations. The blue and red horizontal lines represent the average MFE of pools A and N6, respectively (Figure 1b). The black lines represent the average MFE value of the top 10 000 most abundant sequences in each selection. The composition and diversity of the starting pool and the target of each selection are indicated below each graph. The data show high folding propensity of the most abundant sequences in each selection experiment. Furthermore, selections from mixed pools are dominated by sequences originating from the N6 pool, even if the starting N6:A ratio is 1:10. (b-d) The relative population of the pool N6 sequences in the mixed pools (●, revealed by restriction enzyme analysis) increased for all three selections, as the fraction of recovered RNA (○) increased with the selection rounds for Oregon Green (OG) aptamers (b), NanoLuc/Ni-NTA aptamers (c), and ATP-reacting RNAs (d). During the OG experiment (b), the beads were changed for each selection round to immobilize OG as a carboxy-OG coupled to TentaGel-NH_2 beads (T), or OG−biocytin bound to streptavidin on agarose (A) or polyacrylamide (P) beads. (d) The incubation times for the ATP-reacting RNAs are indicated above the fraction-recovered data points. (e) Analysis of the ∼1000 most abundant sequence clusters for each in vitro selection experiment. Cumulative fraction of sequences derived from pool N6 in the last round of each selection experiment showing that for NanoLuc/Ni-NTA round 7 (gray squares) and round 9 (black squares), as well as the ATP-reacting selections starting from a high- (solid triangles) and low-diversity pools (open triangles), the population is largely dominated by the N6-derived sequences. In the OG selection (circles) the results were mixed, with the most abundant sequence originating from N6 but many other sequences among the top ∼300 clusters derived from pool A.

accessible 2′ hydroxyls (often in dynamic, single-stranded segments), the structured positions tend to be less acylated. We hypothesized that an increase in the prevalence of double-stranded or otherwise structured regions in the selected pools would reduce the total number of backbone positions available for acylation. We performed a bulk acylation of purified transcripts from the A, N5, and N6 pools, with the expectation that extensive acylation (as opposed to low-level acylation used in RNA structure probing analyzed by reverse transcription) will reveal overall solvent accessibility of the pools when analyzed using high-resolution denaturing PAGE. The acylation would result in a mobility shift proportional to the number of acylation positions per RNA strand (Figure S2a). Pool A showed a larger shift due to a higher degree of modification, resulting from a higher solvent-accessible fraction of RNAs, than the N5 and N6 pools (Figure S2b), further supporting the observation that the monomeric pools are less accessible to backbone acylation and are therefore more compact.

Overall, the computational MFE analysis and four independent experimental methods—native PAGE, S1 nuclease digestion, ethidium intercalation, and bulk acylation—all

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

supported the hypothesis that a selection for the monomeric subpopulation of random sequences yields pools of compact, folded RNAs. These results suggest that the selected subpopulation may be rich in functional RNAs and that the relative frequency of these functional molecules, that is, the phenotypic potential of the compact pools, may be significantly higher than the random-sequence pool. To test this second hypothesis, we turned to in vitro selections for functional RNAs from various pool mixtures.

Enrichment of Monomeric Sequences Increases the Phenotypic Potential of the Selected Pools. We started by performing an in vitro selection of aptamers that bind the firefly luciferase (Fluc) protein immobilized on Ni^{2+}-nitrilotriacetic acid (NTA) beads, using only the naïve pool (A). The starting diversity of the DNA template was 5×10^{14}. After 12 rounds of selection, the resulting RNA pool showed detectable binding and elution from Ni-NTA beads. Analysis of the ten most abundant sequences (Table S1) revealed a motif that bound Ni-NTA, eluted with imidazole, and was similar to a previously published Ni^{2+}-binding motif (N11 and N23).[36] This motif, a stem-loop with a conserved CAAUUGN-RAAAACG loop sequence, was present in the most enriched sequence, as well as seven other of the ten most abundant sequence clusters (defined as sets of sequences related by fewer than 10 mutations) isolated in this in vitro selection experiment. Further description of this motif is presented in the NanoLuc aptamer selection below. In silico analysis of the sequenced pool revealed that the two most abundant sequences had MFEs well above the starting pool (A) and that the average MFE for the selected sequences was also above the average MFE for pool A, but lower than the average MFE of pool N6 (Figure 3a). Moreover, six of the ten most abundant clusters were predicted by QGRS Mapper[37] to potentially form G-quartet structures (Table S1), which would provide further stability to the RNAs. The experiment supported the hypothesis guiding this project: that in vitro selection for functional RNAs shifts the population toward more structured sequences. We next turned to selections with mixed pools A and N to test whether enrichment for structured monomers also increased the phenotypic potential of the selected pools.

To explore the phenotypic capacity of the structured pools, we performed in vitro selections against different targets, with the pool N6 directly competing with the pool A for a selectable function. First, we performed an in vitro selection for aptamers binding the fluorescent dye Oregon Green (OG). Pools A and N6 were transcribed at 5×10^{15} and 5×10^{14} diversity, respectively. The pool A diversity was set to the same level as the starting diversity for the monomer selection described above; these diversities therefore represent the maximum practical diversity for pool A and maximum theoretical diversity for pool N6. The PAGE-purified transcripts from both pools were combined, and a selection for OG aptamers was performed for seven rounds, alternating the support media among agarose, polyacrylamide, and Tentagel to prevent isolation of bead-binding aptamers. After 7 rounds, the resulting pool was analyzed by HTS, revealing that the first and the third most frequent sequences originated from pool N6, with one other N6-derived sequence appearing in the ten most abundant clusters (Supplementary Table 1). The analysis further showed that these top clusters consisted of highly structured RNAs, with MFEs well above the average MFE for the entire selected OG pool, as well as above the average MFEs of the two starting pools A and N6 (Figures 1b and 3a). Apta-Seq analysis,[38] a method based on 2′ hydroxyl acylation detected by reverse transcriptase termination mapping[35] of the selected pools in the presence of varying concentrations of the target ligand (Figure S3), revealed that the three most abundant sequences were likely highly structured, although the analysis yielded quantitative information about the strength of the aptamer-ligand interactions only for the most abundant sequence. Overall, HTS revealed that sequences from the structured pool (N6) represented approximately 60% of the selected population of RNAs. In contrast, restriction enzyme analysis of the relative composition of the individual selection rounds showed a somewhat lower fraction of pool-N–derived sequences (~47%; Figure 3b), suggesting that even after seven rounds, the pool contained a large number of pool-A–derived sequences that were not sampled by HTS. Considering that the starting population of the OG aptamer selection was 10-fold biased toward pool A, the final distribution of the two pools was an indication that pool-N6–derived sequences were about 10–15-fold enriched for OG aptamers. This result further indicated that in order for the top sequences to be selected, they had to be generally more structured than the sequences found in pool A, and implied a higher phenotypic potential of the structured pool N6.

Next, we performed a selection against Ni-NTA–immobilized Nanoluciferase (NanoLuc). In this case we chose to combine pools A and N6 at the same diversity of 5×10^{14}, matching the maximum theoretical complexity of pool N6. Even though the selection started with equal parts of the two pools, the population after just one round was >80% skewed toward pool A (Figure 3c), perhaps because the RNAs carry forward many aggregation-prone pool-A-derived sequences, which are presumably nonfunctional. It then took another five rounds of selection to equalize the fractions of the two starting pools, at which point the binding became evident, and the population started to be dominated by a small number of N6-derived sequences. As the selection progressed and the fraction of pool N6 in the total DNA population steadily increased, N6 sequences began dominating the whole population, not just the most abundant sequences (Figure 3c). After 7 rounds, 9 out of the 10 most frequent sequences originated from pool N6, whereas only 1 originated from pool A (Figure 3a). The pool-A-derived sequence was the fourth most frequent, but represented only about 1/10 of the three most abundant, N6-derived sequences. Similarly to what was observed in the OG selection, the majority of the top-ten sequences had a predicted MFE either equal to, or greater than, the average MFE of the pool N6 (Figure 3a).

Sequence analysis of the NanoLuc/Ni-NTA pool revealed that the conserved RNA motif found in the Fluc/Ni-NTA selection was present in 6 of the most abundant 8 sequence clusters (Figure 4a,b), notably, in the top 3 sequences that dominated the selection. This motif consists of a sequence-conserved (AAUUGNRAAAAC) loop flanked by a stem (P1) with a C-G base-pair closing the loop. The stem also often contains a single bulged A in the upstream strand, one or two base-pairs below the conserved loop (Figure 4b). As in the Fluc/Ni-NTA experiment, this motif maps next to different sequences in each of the six dominant clusters, implying that these six sequence clusters emerged independently. The P1 composition is diverse (except where it overlaps with the primer-binding sequences), exhibiting strong covariation and minimal sequence conservation (Figure 4a); therefore, the

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX–XXX

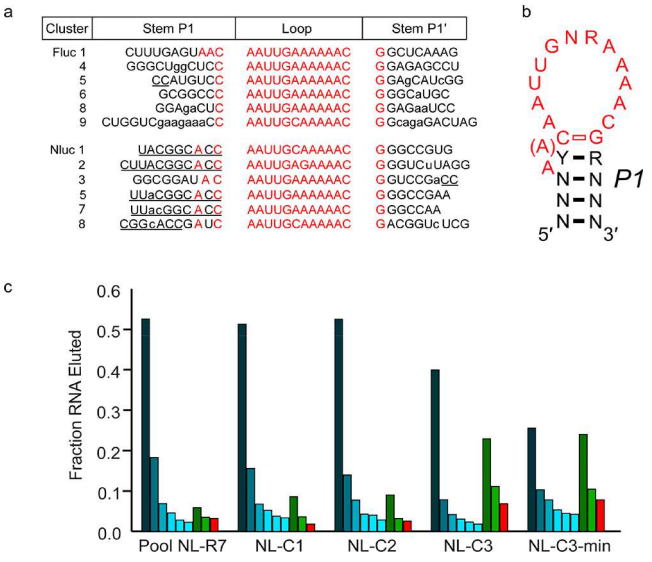


Figure 4. Highly abundant aptamer motif from the Ni-NTA selections. (a) Sequence alignment of a motif found in both Fluc/Ni-NTA and NanoLuc/Ni-NTA aptamer selections. Residues from the constant regions are underlined, conserved residues are highlighted in red, and base-pair mismatches are shown in lowercase. The clusters depict a consensus sequence and conserved structural motif (b) consisting of a stem with potentially one bulging adenosine and a sequence-conserved loop, highlighted in red. The proposed closing C-G base-pair is conserved and therefore not supported by sequence covariation. The bulging adenosine is observed after the Y-R base pair or immediately after the conserved C-G base pair. (c) Column binding profile of the round-7 NL/Ni-NTA pool, the top three NL clusters (C1−C3), and a minimal motif derived from cluster 3. Fractions corresponding to flow-through and washes are indicated in blue, imidazole-eluted fractions in green, and RNA retained on resin in red. Cluster 3 shows the highest binding to the Ni-NTA beads and the NL-C3 min motif derived from this cluster exhibits similar binding and elution profile.

helix likely serves only a structural role. To confirm the activity of the motif, we performed a column-binding assay with the Ni-NTA beads. The top three sequences (NL-C1 through C3), as well as a minimized construct corresponding to the identified motif (NL-C3 min), bound the Ni-NTA beads and eluted with imidazole (Figure 4c). Control experiments with the minimal motif (C3 min) showed similar binding and elution when presented with His_6-tagged proteins (NanoLuc, T7 RNA polymerase, Fluc, and PARK7) and protein-free Ni-NTA beads. Furthermore, the motif did not exhibit binding to the proteins in gel-shift and filter-binding assays, confirming that the binding target of the motif is the chelated Ni^{2+} and not the proteins.

The selection of the same Ni-NTA−binding motif in both the Fluc and NanoLuc selections allowed us to directly compare the theoretical stability of the individual motifs in the most abundant clusters. The motifs derived from the two selections had similar average MFEs, but the distribution of the predicted stabilities was wider for the sequences derived from pool A than pool N6, providing additional evidence that the selection for compact monomeric sequences yields motifs that are on average more structured.

Because the NanoLuc/Ni-NTA selection was dominated by the Ni^{2+}-binding motifs, we continued the experiment by immobilizing the NanoLuc protein on nitrocellulose membrane and isolating RNAs that remained attached to the membrane after extensive washes. After two rounds of selection the abundance of the Ni^{2+}-binding motif started to decrease, none of the sequences among the 10 most abundant clusters were derived from pool A, and the MFE analysis showed that the clusters formed well-folded RNAs (Figure 3a). The top pool-A−derived cluster was the 14th most abundant, representing under 3% of the cumulative sequences up to that point in the ranking. Pool-N−derived sequences heavily dominated the top 1000 most abundant sequence clusters of the selection (Figure 3e). These results show that the pool-A−derived sequences are present in the selected population, but not as strongly enriched as pool N6 sequences, supporting the conjecture that many pool A sequences are aggregating "free-riders" and that the N6 pool has a higher phenotypic potential.

Catalytic Potential of Pools A and N6. To test the relative catalytic potential of the A and N6 pools, we performed an in vitro selection for ATP reactivity, following previously described selections that yielded 5′ self-capping ribozymes.[39,40] This experiment was designed to test the phenotypic capacity of the pools near the lower end of the theoretical diversity of the pool N6 (Figure 1c and Table 1). ATP-agarose beads were incubated with RNAs transcribed from DNA libraries of 5×10^{12} molecules from both pools A and N6. Similarly to the aptamer selections, purified RNAs from the two pools were mixed together at the beginning of the in vitro selection in order for the two pools to compete with each other. After incubation with ATP-agarose beads, 8 washes using two different denaturing buffers were performed to remove noncovalently bound RNAs. On-bead reverse transcription was followed by amplification, and after 8 rounds of selection, the pools exhibited detectable ATP conjugation (Figure 3d). Restriction enzyme analysis showed that in this selection, RNAs from the N6 pool started to dominate the selected population earlier than in the aptamer selections (Figures 3b−d), essentially taking over the population by the third round. HTS of the last round of the selection revealed that the ten most abundant sequence clusters originated from pool N6 and the dominant sequences were more stable than the pool N6 average (Figure 3a). Testing the top four sequences for reactivity toward immobilized ATP confirmed that these are functional RNAs (Figure 5c). The four most abundant sequences share a structural motif (a hairpin with a (A)CAA/AMV(C) internal loop, where M can be either A or C and V is any nucleotide but U; flanked by a CGC/GCG inside helix; Figure 5a,b), suggesting that this motif is important for the ATP reactivity.

Given the limited diversity of the starting population and the denaturing washes during each round of enrichment, the results suggest that the selection quickly suppressed aggregation-prone "free riders" originating from pool A, and revealed the high relative abundance of functional RNAs in the pool N6. Indeed, only 0.12% (4 out of 3340) of the sequences that showed enrichment by having at least two copies in the HTS output originated from pool A. These four sequences mapped to two distinct clusters of 2 sequences each, representing the smallest detectable enrichment. This observation indicated that the abundance of ATP-reacting sequences in the 5×10^{12} sequences of pool A is about 10^3-times lower than in pool N6, and suggested that pool N6 has much higher phenotypic potential.

To compare the catalytic potential of the two pools at the upper end of their complexities, we again combined pools A and N6 at 5×10^{15} and 5×10^{14} diversity, respectively, as we

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

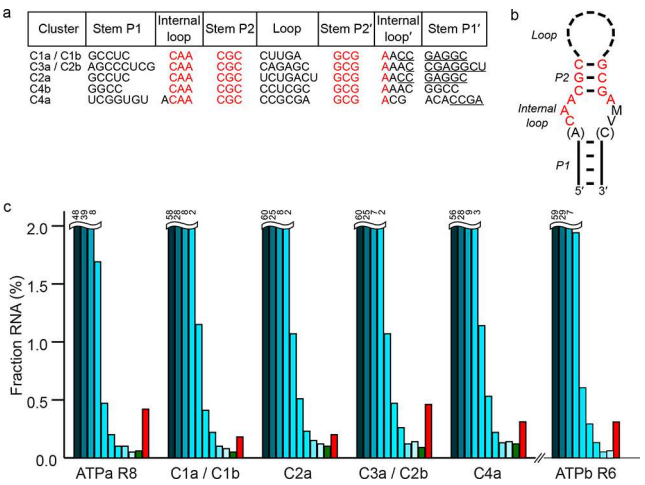


Figure 5. Characterization of dominant sequences form the in vitro selections for ATP-reacting RNAs. (a) Sequence alignment of a motif found in both low- and high-diversity ATP-reacting selections. Residues from the 3′ constant region of the pools are underlined. The clusters depict a conserved structural motif (b) consisting of a stem with an internal loop and a nonconserved loop. Conserved residues are highlighted in red. (c) ATP-column-binding profile of the selected pools (ATPaR8, low-diversity selection, round 8; ATPbR6, high-diversity selection, round 6) and four of the most abundant clusters that contain the conserved motif. Fractions corresponding to flow-through and washes are indicated in blue, ATP elutions are shown in green (when present), and RNA retained on ATP-agarose beads is shown in red. The graphs were truncated at 2%, with actual values indicated above each bar.

had done for the OG selection. We repeated the in vitro selection for reactivity with ATP and found that the dominant sequence clusters were again derived from pool N6 (Figure 5a,5b), with the two most abundant sequences matching the first and the third most abundant sequence from the previous, low-complexity selection. The appearance of the two sequences may have resulted from a contamination by the previous selection; however, none of the other high-abundance sequences from the low-complexity selection were detected in this selection, suggesting that the top two sequences (both originating from pool N6) were isolated independently. The fourth most abundant sequence (also from pool N6) contained the above-described internal loop motif, further suggesting that this motif is important for ATP reactivity. The selected pool was again heavily dominated by sequence clusters derived from pool N6, with pool-A-derived sequences populating low-abundance sequence clusters (Figure 3e). Overall, the results from these two in vitro selections reinforced the conclusion reached in the aptamer selections: that pre-enrichment for monomers leads to an increase in the phenotypic potential of a random-sequence pool.

DISCUSSION

Functional RNAs, such as aptamers and ribozymes, are found in many biological settings and have been evolved in vitro for a variety of roles, for example as regulatory RNAs, functional units of diagnostic and therapeutic tools, and paradigmatic catalysts in Origin of Life research. A hallmark of these RNAs is the formation of specific, compact structures that give rise to their biochemical activities.[41,42] In most known cases, functional RNAs form partially solvent-inaccessible domains, which are often associated with the binding and catalytic sites, and solvent-exposed peripheral segments that typically serve as topological connections or tertiary contacts.[2,41] In line with these observations, our computational analysis suggested that RNA pools containing functional motifs have increased secondary-structure content when compared to random-sequence RNA pools (Figures 1a and S1a) and in vitro selections for aptamers from a fully random pool yielded sequences with increased secondary-structure content and associated stability (Figures 3a and 4a,b). The emergence of functional RNAs, whether in vivo or in vitro, thus coincides with the formation of compact structures. While the structures of cellular RNAs benefit from the long process of evolutionary fine-tuning, RNAs evolved in the laboratory typically undergo a very short selection and must therefore possess a functional global fold and at least partial biochemical activity from the start of the experiment. Because at least some activity has to be present in the founding population, the phenotypic potential, the ability of a sequence pool to perform a selectable function, of the initial pool is critical for the success of the selection.

Generally, the phenotypic potential scales with the molecular diversity of the population; however, random-sequence pools also tend to aggregate, leading to potential sequestration of active sequences within large aggregates or to propagation of "free-riders" that bind the functional sequences. The conundrum presented by pool aggregation and the need for compact folding of functional RNAs led us to explore an experimental approach aimed at increasing the prevalence of structures within a random-sequence pool and testing the phenotypic potential of these structured pools. Purification of monomeric sequences indeed increased the prevalence of sequences that were not only resistant to aggregation with random sequences, but were compactly folded as well (Figures 1c and S1b). HTS analysis showed that their average stability increased with the native-PAGE selection round, reaching a plateau after 5 rounds of selection, and matching the folding propensities of the in silico generated pools (Figure 1b). Two concomitant effects appeared to take place: elimination of sequences with low predicted secondary-structure content (low conformational stability), which likely directly resulted in the low aggregation propensity of the selected pools, and shifting of the population toward sequences with stable, structured domains resistant to enzymatic digestion (Figure 2a) and increase of structure and compaction (Figures 2b and S2). The experimental data were thus in accordance with the HTS analysis, showing that isolation of the nonaggregating, monomeric subpopulation of a random-sequence pool yields structured, compactly folded sequences. The results also suggested that further selection of the monomeric subpopulation of the sequences was not likely to increase the folded fraction (Figures 1b and 2) or perhaps even the structural diversity of the selected RNAs. On the other hand, further mutagenesis to explore the sequence space around the structured sequences may provide an avenue for additional increase in the functional diversity of the compact pools. Possible scenarios for this functional increase include (1) acquisition of new functions based on the same overall fold,[16,43,44] (2) exploration of bistable motifs allowing attainment of both a new fold and a new function,[45−47] and (3) discovery of new active motifs. Low-level mutagenesis introduced during the selection of the compact pools (Figure 1d) could, in principle, lead to such innovation; however, our in vitro selections were not designed to test for evidence of these mechanisms.

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Because the compact, monomeric pools had undergone four rounds of transcription, reverse-transcription, and PCR amplification (to yield pools N3−N6), the selection also likely biased the population toward more amplifiable sequences. It is difficult to assess the contribution of the three polymerases to this bias, although some basic observations can be made. For example, we expect that sequences or structures that lead to transcription termination are eliminated from the population, but these can still aggregate with other sequences when transcribed from a naive pool. However, given their specific structures and sequences,[48] transcription terminators would appear at a very low rate (<0.1%) in a random pool and would not be propagated if the full-length RNA pool is purified. We therefore do not expect terminators to affect its overall phenotypic potential. Similarly, structures that are difficult to read by certain reverse transcriptases are suppressed during the selection process[16] but are likely present among the initial random-sequence transcripts. Our measurement of primer extension during the first RT step (to make pool N3) showed 96% efficiency, suggesting that the RT step is not a major source of diversity loss. Finally, homopolymeric sequences are also likely suppressed either because they form structures that are too stable to unfold (such as G-quadruplexes) or potentially not form compact, stable structures (such as poly-A or poly-U sequences). Clearly, some sequence and structural bias must be introduced during the selection cycles. Our analysis revealed that structured pools are readily amplified, but it remains to be established what folds, if any, are more likely to dominate the selections.

Previous experiments with equimolar mixtures of fully random and stem-loop-containing RNA pools revealed that the hairpin-containing pool yielded aptamers with more structural modules and significantly higher activity.[15,49] We hypothesize here that selection for structured RNAs leads to an increase of the phenotypic potential of the RNA pool and we also performed competitive selection experiments, but the pools were mixed in varying proportions at the beginning of each selection. RNAs from the compact pool dominated each selection, even when the sequence diversity of the starting pool was biased against the compact pool (Figure 3a). Our results consistently showed that compact, monomeric sequences derived from random-sequence pools are biochemically rich, with a significantly higher phenotypic potential than the parent, fully random pool.

A quantitative measure of the relative phenotypic potential of the two pools was revealed by a ribozyme selection that started with low pool diversity, yielding functional sequences completely dominated by the compact-pool sequences (Figures 3a,d,e and 5 and Table S1). These results imply that the compact pool has at least 1000-times higher frequency of functional RNAs than pool A, meaning that 5×10^{12} sequences of the compact pool N6 yield about as many functional RNAs as 5×10^{15} sequences of the fully random pool A. Given the near-zero enrichment of pool-A sequences, the phenotypic potential is likely more than 1000-times higher in pool N6 than in pool A. This observation suggests that the original selection for monomeric sequences largely preserved the functional RNAs present in the original pool (A), because the selection decreased the number of pool-A sequences from 5×10^{15} to $\sim 5 \times 10^{11}$ (Table 1 and Figure 1d). Assuming unbiased propagation of sequences, pool N6 should be enriched for structured, functional RNAs by at least 10^4 times and the selection for ATP-reacting RNAs suggests that this enhancement is nearly mirrored in the phenotypic potential as well. The selection of compact, monomeric sequences thus, for the first time, represents an experimental approach to greatly increase the abundance of functional RNAs accessible by in vitro selections. We expect that this phenotypic enhancement would take place even if the diversity of the starting random-sequence pool is lower than 5×10^{15}.

CONCLUSIONS

Our results show that a preparatory-scale purification can be used to enhance the phenotypic potential of a random RNA pool. By purifying RNAs transcribed from a random DNA pool using a native-PAGE fractionation, we isolated sequences that are not only more structured but also enriched for functional RNAs. This method increases the phenotypic potential of random-sequence pools by several orders of magnitude and allows lengthening of the random-sequence regions to explore more complex structures, which can give rise to higher biochemical activity.[23,49] While the immediate application of the preparatory technique is to increase the phenotypic potential of random-sequence RNA pools in order to improve the outcomes of in vitro selection experiments, we note that the removal of aggregating RNAs could also be relevant for the Origin of Life and the evolution of early biosphere. Removal of aggregating RNAs molecules, whether through mineral-mediated molecular filtration, surface adsorption, or simple precipitation, increases the prevalence of stably folded sequences with higher phenotypic potential, reducing the sequence space that needs to be explored in order to find functional folds. This reduction could have implications for the emergence of the first catalytic RNAs. For example, it has been shown that a medium-sized ribozyme such as the class I ligase, with a size similar to that of the RNA pools discussed in this work, would appear only once in 3×10^{18} molecules of a random RNA sequence[23] and simpler functional motifs are likely far more common.[11] A physicochemical process that removes aggregating sequences may significantly lower the requirement of sequence diversity to yield functional molecules, thereby increasing the phenotypic potential of the emerging sequences of the RNA world. Furthermore, any process that introduces sequence diversity to such a population through recombination or single-nucleotide mutagenesis would also likely increase the phenotypic potential of the RNAs.

Finally, the enhancement of phenotypic potential of random-sequence pools brings closer the advent of de novo discovery of functional RNAs through screening of individual sequences. For example, given the speed of detection using ultrafast cell-sorters,[50] at $\sim 10^5$ events/s, it is feasible to sort $\sim 10^{10}$ droplets or beads in a day. While exploring the activity of 10^{15} sequences would be impractical, even if 100−1000 random RNAs are loaded into each bead or droplet, it may be feasible to screen for a biochemical function by detecting the signal generated by individual sequences in an entire compact pool, such as pool N6.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.9b11396.

Supporting text, including Materials and Methods, supporting Figures S1−S3, supporting references (PDF)

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

Supplementary Table S1 containing designed and selected DNA sequences, HTS statistics, MFE calculations for the dominant sequence clusters, and G-quartet predictions (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Jack W. Szostak − *Massachusetts General Hospital, Boston, Massachusetts, and Harvard University, Cambridge, Massachusetts;* orcid.org/0000-0003-4131-1203; Email: szostak@molbio.mgh.harvard.edu

Andrej Lupták − *University of California at Irvine, Irvine, California;* orcid.org/0000-0002-0632-5442; Email: aluptak@uci.edu

Other Authors

Fabio Chizzolini − *University of California at Irvine, Irvine, California;* orcid.org/0000-0003-4455-8367

Luiz F. M. Passalacqua − *University of California at Irvine, Irvine, California;* orcid.org/0000-0002-5490-2427

Mona Oumais − *University of California at Irvine, Irvine, California;* orcid.org/0000-0002-5327-1379

Armine I. Dingilian − *University of California at Irvine, Irvine, California;* orcid.org/0000-0003-3572-2014

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.9b11396

Author Contributions

#F.C., L.F.M.P., and M.O. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the John Templeton Foundation through the Foundation for Applied Molecular Evolution and NSF-CBET 1804220 (A. L.), NSF-IGERT Biophotonics predoctoral fellowship (M. O.), and Science Without Borders Program, CAPES Foundation, Ministry of Education of Brazil, Process 99999.013571/2013-03 and Miguel Velez Scholarship-UCI (L. P.). The authors thank the following laboratories for their generous contributions: Spitale lab (UCI) for the acylating reagents, the Prescher lab (UCI) for purified firefly and NanoLuc luciferases, and Weiss lab (UCI) for purified PARK7 protein. The authors also thank K. A. Dunne-Dombrink for help with the Fluc/Ni-NTA selection. J.W.S. is an investigator of the Howard Hughes Medical Institute.

REFERENCES

(1) Hermann, T.; Patel, D. J. Stitching Together RNA Tertiary Architectures. *J. Mol. Biol.* **1999**, *294* (4), 829−849.

(2) Edwards, T. E.; Klein, D. J.; Ferré-D'Amaré, A. R. Riboswitches: Small-Molecule Recognition by Gene Regulatory RNAs. *Curr. Opin. Struct. Biol.* **2007**, *17* (3), 273−279.

(3) Serganov, A.; Patel, D. J. Ribozymes, Riboswitches and beyond: Regulation of Gene Expression without Proteins. *Nat. Rev. Genet.* **2007**, *8* (10), 776−790.

(4) Jimenez, R. M.; Polanco, J. A.; Lupták, A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem. Sci.* **2015**, *40* (11), 648−661.

(5) Russell, R.; Millett, I. S.; Tate, M. W.; Kwok, L. W.; Nakatani, B.; Gruner, S. M.; Mochrie, S. G. J.; Pande, V.; Doniach, S.; Herschlag, D.; Pollack, L. Rapid Compaction During RNA Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (7), 4266−4271.

(6) Woodson, S. A. Compact Intermediates in RNA Folding. *Annu. Rev. Biophys.* **2010**, *39* (1), 61−77.

(7) Wilson, D. S.; Szostak, J. W. In Vitro Selection of Functional Nucleic Acids. *Annu. Rev. Biochem.* **1999**, *68*, 611−647.

(8) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346* (6287), 818−822.

(9) Tuerk, C.; Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **1990**, *249* (4968), 505−510.

(10) Robertson, D. L.; Joyce, G. F. Selection in Vitro of an RNA Enzyme That Specifically Cleaves Single-Stranded DNA. *Nature* **1990**, *344* (6265), 467−468.

(11) Pobanz, K.; Lupták, A. Improving the Odds: Influence of Starting Pools on in Vitro Selection Outcomes. *Methods* **2016**, *106*, 14−20.

(12) Schultes, E. A.; Spasic, A.; Mohanty, U.; Bartel, D. P. Compact and Ordered Collapse of Randomly Generated RNA Sequences. *Nat. Struct. Mol. Biol.* **2005**, *12* (12), 1130−1136.

(13) Sabeti, P. C.; Unrau, P. J.; Bartel, D. P. Accessing Rare Activities from Random RNA Sequences: The Importance of the Length of Molecules in the Starting Pool. *Chem. Biol.* **1997**, *4* (10), 767−774.

(14) Legiewicz, M.; Lozupone, C.; Knight, R.; Yarus, M. Size, Constant Sequences, and Optimal Selection. *RNA* **2005**, *11* (11), 1701−1709.

(15) Davis, J. H.; Szostak, J. W. Isolation of High-Affinity GTP Aptamers from Partially Structured RNA Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11616−11621.

(16) Porter, E. B.; Polaski, J. T.; Morck, M. M.; Batey, R. T. Recurrent RNA Motifs as Scaffolds for Genetically Encodable Small-Molecule Biosensors. *Nat. Chem. Biol.* **2017**, *13* (3), 295−301.

(17) Ikawa, Y.; Tsuda, K.; Matsumura, S.; Inoue, T. De Novo Synthesis and Development of an RNA Enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (38), 13750−13755.

(18) Ruff, K. M.; Snyder, T. M.; Liu, D. R. Enhanced Functional Potential of Nucleic Acid Aptamer Libraries Patterned to Increase Secondary Structure. *J. Am. Chem. Soc.* **2010**, *132* (27), 9453−9464.

(19) Luo, X.; Mckeague, M.; Pitre, S.; Dumontier, M.; Green, J.; Golshani, A.; Derosa, M. C.; Dehne, F. Computational Approaches toward the Design of Pools for the in Vitro Selection of Complex Aptamers. *RNA* **2010**, *16* (11), 2252−2262.

(20) Sassanfar, M.; Szostak, J. W. An RNA Motif That Binds ATP. *Nature* **1993**, *364* (6437), 550−553.

(21) Carothers, J. M.; Davis, J. H.; Chou, J. J.; Szostak, J. W. Solution Structure of an Informationally Complex High-Affinity RNA Aptamer to GTP. *RNA* **2006**, *12* (4), 567−579.

(22) Grate, D.; Wilson, C. Laser-Mediated, Site-Specific Inactivation of RNA Transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (11), 6131−6136.

(23) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences. *Science* **1995**, *269* (5222), 364−370.

(24) Illangasekare, M.; Sanchez, G.; Nickles, T.; Yarus, M. Aminoacyl-RNA Synthesis Catalyzed by an RNA. *Science* **1995**, *267* (5198), 643−647.

(25) Curtis, E. A.; Bartel, D. P. New Catalytic Structures from an Existing Ribozyme. *Nat. Struct. Mol. Biol.* **2005**, *12* (11), 994−1000.

(26) Rogers, J.; Joyce, G. F. The Effect of Cytidine on the Structure and Function of an RNA Ligase Ribozyme. *RNA* **2001**, *7* (3), 395−404.

(27) Robertson, M. P.; Ellington, A. D. In Vitro Selection of an Allosteric Ribozyme That Transduces Analytes to Amplicons. *Nat. Biotechnol.* **1999**, *17* (1), 62−66.

(28) Robertson, M. P.; Scott, W. G. The Structural Basis of Ribozyme-Catalyzed RNA Assembly. *Science* **2007**, *315* (5818), 1549−1553.
(29) Pitt, J. N.; Ferré-D'Amaré, A. R. Structure-Guided Engineering of the Regioselectivity of RNA Ligase Ribozymes. *J. Am. Chem. Soc.* **2009**, *131* (10), 3532−3540.
(30) Cadwell, R. C.; Joyce, G. F. Randomization of Genes by PCR Mutagenesis. *Genome Res.* **1992**, *2* (1), 28−33.
(31) Wilson, D. S.; Keefe, A. D. Random Mutagenesis by PCR. *Current Protocols in Molecular Biology* **2001**, 8-3-1−8-3-9.
(32) Duhamel, J.; Liu, D. M.; Evilia, C.; Fleysh, N.; Dinter-Gottlieb, G.; Lu, P. Secondary Structure Content of the HDV Ribozyme in 95% Formamide. *Nucleic Acids Res.* **1996**, *24* (20), 3911−7.
(33) Vogt, V. M. Purification and Further Properties of Single-Strand-Specific Nuclease from Aspergillus Oryzae. *Eur. J. Biochem.* **1973**, *33* (1), 192−200.
(34) Lepecq, J. B.; Paoletti, C. A Fluorescent Complex between Ethidium Bromide and Nucleic Acids. Physical-Chemical Characterization. *J. Mol. Biol.* **1967**, *27* (1), 87−106.
(35) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **2005**, *127* (12), 4223−4231.
(36) Hofmann, H. P.; Limmer, S.; Hornung, V.; Sprinzl, M. Ni2+-Binding RNA Motifs with an Asymmetric Purine-Rich Internal Loop and a G-A Base Pair. *RNA* **1997**, *3* (11), 1289−1300.
(37) Kikin, O.; D'Antonio, L.; Bagga, P. S. QGRS Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Res.* **2006**, *34*, W676−W682.
(38) Abdelsayed, M. M.; Ho, B. T.; Vu, M. M. K.; Polanco, J.; Spitale, R. C.; Lupták, A. Multiplex Aptamer Discovery through Apta-Seq and Its Application to ATP Aptamers Derived from Human-Genomic SELEX. *ACS Chem. Biol.* **2017**, *12* (8), 2149−2156.
(39) Zaher, H. S.; Watkins, R. A.; Unrau, P. J. Two Independently Selected Capping Ribozymes Share Similar Substrate Requirements. *RNA* **2006**, *12* (11), 1949−1958.
(40) Huang, F.; Yarus, M. 5'-RNA Self-Capping from Guanosine Diphosphate. *Biochemistry* **1997**, *36* (22), 6557−6563.
(41) Ferré-D'Amaré, A. R.; Doudna, J. A. RNA Folds: Insights from Recent Crystal Structures. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28* (1), 57−73.
(42) Alemán, E. A.; Lamichhane, R.; Rueda, D. Exploring RNA Folding One Molecule at a Time. *Curr. Opin. Chem. Biol.* **2008**, *12* (6), 647−654.
(43) Kim, J. N.; Roth, A.; Breaker, R. R. Guanine Riboswitch Variants from Mesoplasma Florum Selectively Recognize 2 '-Deoxyguanosine. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (41), 16092−16097.
(44) Huang, Z.; Szostak, J. W. Evolution of Aptamers with a New Specificity and New Secondary Structures from an ATP Aptamer. *RNA* **2003**, *9* (12), 1456−1463.
(45) Schultes, E. A.; Bartel, D. P. One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds. *Science* **2000**, *289* (5478), 448−452.
(46) Bendixsen, D. P.; Collet, J.; Østman, B.; Hayden, E. J. Genotype Network Intersections Promote Evolutionary Innovation. *PLoS Biol.* **2019**, *17* (5), No. e3000300.
(47) Held, D. M.; Greathouse, S. T.; Agrawal, A.; Burke, D. H. Evolutionary Landscapes for the Acquisition of New Ligand Recognition by RNA Aptamers. *J. Mol. Evol.* **2003**, *57* (3), 299−308.
(48) Lyakhov, D. L.; He, B.; Zhang, X.; Studier, F. W.; Dunn, J. J.; McAllister, W. T. Pausing and Termination by Bacteriophage T7 RNA Polymerase. *J. Mol. Biol.* **1998**, *280* (2), 201−213.
(49) Carothers, J. M.; Oestreich, S. C.; Davis, J. H.; Szostak, J. W. Informational Complexity and Functional Activity of RNA Structures. *J. Am. Chem. Soc.* **2004**, *126* (16), 5130−5137.
(50) Lau, A. K. S.; Wong, T. T. W.; Shum, H. C.; Wong, K. K. Y.; Tsia, K. K. Ultrafast Microfluidic Cellular Imaging by Optical Time-Stretch. In *Methods Mol. Biol.*; Humana Press: New York, 2016; Vol. *1389*, pp 23−45.

https://dx.doi.org/10.1021/jacs.9b11396
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX